DOCUMENT RESUME

ED 076 498 SP 006 254

AUTHOR Cluff, Robert B.; Smith, Charles W.
TITLE Humanizing the Testing Aspect of Evaluation.
NOTE 27p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Pretests; *Student Testing; Teacher Education; Teaching Techniques; *Test Construction; *Testing; *Test Reliability; Test Results; *Test Selection

ABSTRACT

This paper presents guidelines for valid and effective test preparation and presentation. Working on the assumption that teacher secrecy about tests and the subsequent student test-fear are pointless and harmful, this paper outlines some tools for test preparation, such as pretests, teacher file cards as to items covered in class and clearly outlined course objectives. Also covered are some principles to be considered in preparing tests--such as, for a test to be valid, questions must be in line with what was actually taught. The paper concludes that tests must be more humanly conceived, but they still must not be thought of as ends in themselves. (JA)

FILMED FROM BEST AVAILABLE COPY

ED 076498

140

ROBERT B. CLUFF and CHARLES W. SMITH

Northern Illinois University, DeKalb

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Humanizing the Testing Aspect of Evaluation

Case Study

Paul sits awaiting the test promised for the next class. He wonders about the nature of the examination. Will it be objective as was the previous test? As time passes, Paul's anxiety increases. He anticipates the probable result if perchance the test should be essay. The teacher's philosophy is that the student should be prepared for any eventuality.

SP 006 354

The test time has arrived as has the teacher. In short order the tests are distributed. Due to the length of the test, no questions are entertained at the beginning of the class--to do so might crowd the testing time. As Paul examines the test he realizes his fears to be correct. The test, previously veiled in secrecy, is now revealed to be a five-question essay exam from which Paul is to respond to any three items. Paul becomes increasingly nervous and anxietous as he analyzes the test. Many questions come to mind:

What was the teacher's purpose and rationale for choosing the essay exam?

Will my preparation be beneficial inasmuch as the test is so different from the type for which I studied?

How much time shall I spend on each of the essay items? I wish the directions were more explicit. In

some previous experiences of this nature I have run out of time only to find that I had neglected to complete those items that the teacher believed to be most important.

How will the examination be scored?

What use will be made of the results?

Case Study: Initial Comments

Paul's feelings are not uncommon in situations where the above questions remain unanswered. It is probable and likely that Paul is using so much energy and time dealing with his test anxiety that his test results will not realistically reflect his knowledge of the subject.

The teacher in the case study above could have done much to humanize his testing procedures. Lessened anxiety and a more favorable attitude toward testing would have resulted if the learner had been made aware of the purpose of the

test, guided toward adequate preparation, provided with specific directions during test administration, informed of scoring procedures and notified of the use to be made of the test results.

In part, test fear may be allayed through frequent quizzes that may or may not count toward a grade. Such quizzes are helpful to students in preparing for later examinations. They emphasize what the teacher feels to be important learning outcomes. The value of instant tests should not be limited to the alleviation of student fears. Such tests provide valuable insight to the teacher regarding instructional strengths and weaknesses and to the student by answering his questions of, "What do I know?" and "What do I need to know?"

Tests With a Purpose

Obviously testing must be closely related to the overall instructional process. A pretest may serve to familiarize students with an

overview of the course. It may be helpful to students by indicating the type of examination items that a particular teacher utilizes. In addition, test results from the pretest may provide information for grouping students by achievement or readiness level or it may indicate a starting point for instruction. Pretest results will also serve to indicate the extent of growth realized when compared with post-test measures at later phases of instruction. It should be noted that the test purpose should be positive. Testing should never be done as a punitive measure. Such use may cause students to dislike teachers, testing, school, learning, and related activities.

Teacher verbal discourse and test item construction have been legitimately criticized as consisting predominantly of unrelated facts. This lowest level of thinking (1), though important and necessary, should not be stressed exclusively in

a teaching program. Desirably a teacher should attempt to structure classroom questions (2) and test items based upon various levels of the cognitive domain. Thus, emphasis in testing should reflect higher aspects of thinking, comprehension and application, as opposed to merely rote memorization. Knowledge acquired at these higher levels of cognition is more likely to be retained (3) and put to use than are isolated facts. The question is, "What type of testing is most appropriate for measurement of higher learning processes?"

Every type of test item has particular uses and limitations. During test construction, it is imperative that the teacher give special consideration to whether or not he is using the item types that are most appropriate for gaining the desired information. Also, students exhibit varying degrees of success with different types of items.

Consideration should therefore be given to the utilization of various item types in test construction.

Test Preparation

Effective item writing develops, as does any other skill, with practice, evaluation of one's efforts and further practice based upon the feedback received. Item writing is perhaps best facilitated by storing each item on a separate index card. The front of the card should include the item itself with adequate space for penciled corrections, and a description of the item content. The other side of the card should include item analysis data--the difficulty level and the discrimination power. A file should be maintained of items that prove to be most valuable in the learning process. Such a file aids one in retaining items that have proven to be less distressful to students and items that are more revealing in diagnosis. A sizable test item facilitates test

construction and permits quick preparation of alternate forms.

In the testing process the mind becomes conditioned to specific item types. For example, the mind is conditioned to dual response in the true-false and to an entirely different response--recall--for short answer items. To require students to switch frequently from one item type to another is to impose an additional burden in the testing situation. This may not be terribly troublesome to some students but others are bothered to the extent of loss of efficiency in responding.

The achievement test should contain items at various levels of difficulty. Motivation in the testing situation, quite necessary for some learners, may be enhanced by including the easiest items at the beginning of the test. By so arranging the items, students may pick up bits of information that will permit them to form concepts, which in turn

will permit them to answer some of the more difficult items. In this sense testing becomes a learning process and the test becomes a teaching device.

Students legitimately complain of the "unfair" test. For the test to be a valid one, questions should be in line with what has actually been taught. The implemented objectives, instructional materials used, and the methods of presentation should all be reflected in the test. Students should be made aware of the course objectives at the outset of the course. Test items should also evolve from the class text, required outside readings, handouts presented, and resultant class discussions of such assignments. For example, a test consisting of modern math terms, symbols and operations would be an invalid and unfair assessment instrument for a group of students who has been taught traditional arithmetic. It is imperative that test items

reflect the content stressed, methods used and desired objectives of the course. Since reliability and validity are both influenced by the number of items included in the test, the more frequent the testing the greater the number of items affecting the grade. Increasing test length, if not overdone, will also tend to improve reliability and validity.

Preparing the test as the unit progresses, rather than at the end of the unit of instruction, is one way to insure that the items are valid ones for the group being tested. In this way one can be sure of testing what has been taught and not relying on memory of what should have been taught. Fairness to the student requires that only those concepts that have been thoroughly presented in class should be tested. As stated by Bugelski, "Examination items should be based upon material that has been covered during instruction rather

than the teacher's personally devised extrapolations of such material."

The test should be planned so that hopefully all of the group will be able to complete the test within the specified time limits. It really doesn't matter if the best prepared students finish the test in a relatively short time. Even for them the purposes of the test have been fulfilled. What really matters is that even the slowest learner has sufficient opportunity to demonstrate his achievement. Generally, the teacher should be most concerned about those learners who are more likely to finish last since such students are frequently stymied, frustrated, thwarted, or defeated by the test experience, and usually prove to be those who need additional instruction. Poor test results may indicate that the student needs general assistance in improving his reading skills. In addition, such results may suggest that the student needs a different

type of instruction and a different test.

How much and what types of information should students be given prior to an examination? Proper preparation hinges upon the amount and type of information provided by the teacher. No information which would be beneficial to the students, short of actual questions and answers, should be withheld. Students have the right to know what general purpose is to be served by the test administration. Hopefully no test will be administered to a group unless the teacher or the learner has a need for the information that the test may provide. Use of test results is mandatory if testing is to be respected. Among other uses, a test may purposefully be administered to provide information for parental conferences, to diagnose strengths and weaknesses of group members, to facilitate student awareness of needs and progress, and to aid the teacher

in evaluating his instructional program.

When test results are to be utilized for grading purposes, student preparation for the test is most vital. Students should know in advance when they are to be administered tests that may have a major influence upon their grades. Such information is needed far enough in advance to allow for mental conditioning and for proper preparation. Students have a right to know in advance the types of items, the approximate number of items, time allowed, and the content that is expected to be mastered.

Discretion should be used in determining an appropriate date and time for the examination. Generally students are more excited immediately prior to a field trip, pep rally, or a vacation period. Such a time would be inopportune for a testing situation in which test results were

expected to yield reliable indices of pupil achievement.

When the teacher constructs the test, the directions should include such information as point values and time limits, especially for essay items. By so doing, a student is not as likely to be penalized for neglecting some items. Point values and time guidelines enable the student to spend his time more effectively. All factors that will enter into the grading process should be discussed in the directions.

The test instructions should tell students to temporarily omit difficult items, returning to them as time permits. Such a strategy aids students in that they can respond quickly to items that they know and are able to devote the remaining time to more troublesome items. In this manner, slower students who may be unable to complete the examination in the required time will not be missing items that they could probably be

expected to answer. The students should also be cautioned to mark their answers in the correct positions on the answer sheet when items are skipped.

Every test should include general directions as well as directions for specific parts. General directions might discuss information relevant to what materials may be used during the test, where and how to mark the answers, where to do any necessary calculations, time allowed for each part, scoring procedures and identifying data. Time should be allowed at the beginning of every testing situation to clarify any questions regarding test directions.

Test Administration

It is advisable that the teacher be at the testing session to administer the test. Class morale is thereby enhanced and the instructor performs a service. The role he plays in the testing situation should remain that of teacher--not policeman. Frequently

he will be required to answer questions that might prove puzzling to a proctor. In addition, the teacher may gain considerable insight into students and the class as a whole during the test administration. Opportunities to reinforce learning, gain group insight, and gain information for item analysis will be lost if the teacher is absent during testing.

If the purpose of the test is instructional, and if class size is sufficiently small, students might best write on the test itself. When these are hand-scored, the teacher will learn much about the group that can be utilized to further the instructional process. If the tests are machine-scored, similar information will be obtained through an item analysis and careful study of individual responses.

Scoring - Grading

Students should be aware of the ground rules that will be used in

in scoring the examination. Not to know how a test will be scored--whether or not a penalty or correction for guessing is to be imposed--is to give an unfair advantage to some students. Teachers should be aware that personality structure is an important influence and should attempt to neutralize this factor in the directions and scoring. For example, unique personality structures prohibit some from guessing. At the same time, others with a different personality may guess wildly. Other factors also influence test response and some play an even more significant role, i.e., intelligence, achievement, preparation and motivation.

Testing primarily for grading purposes should be done on a frequent basis rather than only at the end of a unit of instruction. The total number of test items should never be the sole factor for determining grading. The highest score on a test

should always be the upper limit of the highest grade given irregardless of the number of items included. Such a strategy makes some provision for error that every teacher must recognize as existing in each test, for example, error stemming from testing subject matter that has not been thoroughly taught.

When grading accompanies testing, the objective type examination is most feasible. In the total testing program, the teacher will likely employ both objective and essay tests. The essay examination has great potential when used properly. The authors prefer to see the essay exam used as an instructional device, but cannot justify its use for grading due to its gross unreliability of scoring. Students and teachers should view the test situation as an opportunity for gaining insight into one's problems and/or successes. This diagnostic function of testing, and its use in fostering an enriched

learning experience, should far surpass the use of essay tests for grading purposes.

Any factors extraneous to the subject matter being tested should be graded separately from the subject material. It is unfair and unwise for a teacher to deduct points from a social studies examination, for example, because of poor spelling, poor handwriting, and inaccurate syntax. This is not to imply that the social studies test cannot be used to gain insight into language usage. However, where grading is concerned, each of these areas should be graded separately. Consideration of factors extraneous to a subject also interferes with test validity.

The argument is often given that it is not fair to students who achieve high scores to allow others to retake a test and possibly achieve a higher score. Is it unfair to grade the top student "A" then grade others "A" after they have retaken the test and

demonstrated that they have achieved as much knowledge, comprehension or skill as the top student? What does it matter if all of the students get "A" if they have all acquired the expected knowledge and skills? It is an ideal teacher who can accomplish his work so well and he should grade accordingly. On the other hand, the teacher who finds that he must fail a large number of students in his class is admitting his own failure. Objections that the bright student who gets "A" is discriminated against if others are given special opportunities to achieve the same grade may be likened to those in the parable who objected to receiving a penny for a day's work when others who came late in the day also received a penny. As long as the same opportunities are extended to all, there can be no question of fairness.

Contemporary Views - Testing, Evaluation

It should be stressed that a

test is not a sacrosanct instrument. No matter how methodically or laboriously prepared, it is not worth the distress of even one student. Test results are never a true indication of what a student knows or doesn't know. Test results may not give an accurate picture of a student's achievement in relation to others. No single set of test results should ever be the final word in evaluation. If a student does poorly on one test, opportunities should be provided for him to demonstrate that he has indeed achieved what is expected either by retaking the test, taking another test or demonstrating his achievement in other ways.

Grading, like testing, should be regarded as only one indication of a student's achievement. There is nothing sacred or infallible about a grade. The teacher should be willing to raise a grade any time a student is able to demonstrate that he has achieved comparably to others who have

been given a higher grade. Since the teacher's job is to help every individual acquire as much knowledge, understanding and skill as possible, what does it matter how often or under what conditions the student demonstrates his achievement?

If testing is going to serve as an enriching learning experience, reinforcement must be a necessary part of the process. An answer key should be made available to students as soon as possible after the test situation. The test key and accompanying explanations may be posted outside the classroom if students are permitted to leave after taking the test or it may be distributed in class after everyone has completed the test. The student's desire to know answers is at a high peak immediately after the examination. Providing answers to students' questions concerning test items may be extremely valuable in furthering learning at such an opportune time.

As soon as test papers are scored, they should be returned to students. For further reinforcement, these papers should be discussed. An item analysis will indicate those questions that appeared to be most troublesome to students. These should be stressed, but opportunity should be provided for students to discuss any item. This is perhaps the best way for the classroom teacher to get feedback regarding the effectiveness of his instruction, individual achievement, group strengths and weaknesses, instructional areas requiring additional emphasis or correction, and for the improvement of viable test items. Students, through class discussion, may contribute many factors that were not apparent to the teacher during test preparation. At the same time, it is probable that much learning is taking place.

It is important that the instructor prepare a written test

key indicating anticipated responses prior to test administration. A well-prepared key is essential to reliable scoring of essay examinations particularly. The essay examination is only as good as its key. Only those concepts and analyses specified in the key should be given credit in an objective sense. Other extraneous material of value written into an essay response may be discussed in class after the test and tested at a later time.

If a student can present convincing support, logical rationale, or source documentation in support of his answer, he should be given credit for the answer in question. It is likely an item, however, that the teacher may wish to revise before using it again.

On-going evaluation is an important part of the teaching process. Testing of various types is an important aspect of such an evaluation program. Continual attempts should

be made by teachers to make the testing and evaluation phases of instruction enriching learning experiences where threat, fear, and anxiety are minimized. This aspect of the teaching process has a great impact on students. Hopefully each teacher will do his share to make it a favorable impact. Bigge has stated, "A teacher's program of evaluation not only governs his students' study habits, their manner of interaction in class, and the number and quality of their learnings; it also largely determines whether they will pursue learning on a memory-, understanding-, or reflection-level."

Each group is somewhat dissimilar. Effective strategies and techniques planned for one group may not be equally effective for another group. Flexibility must therefore be a part of the testing and evaluation program.

Preferably test results should be considered only as one type of evidence that will be used in

conjunction with other evaluative evidence, i.e., class projects, homework, and class participation. Many factors affect test results including emotional level at the time of testing, health conditions of the learner, environmental factors, motivation, and the child's attitude toward the examiner. Therefore, the outcomes of testing are never, in and of themselves, conclusive evidence of a child's achievement. Since evaluation techniques are tools of the teaching profession, the test, a means to an end, must not be construed as, or used as, an end in itself.

References

1. Bloom, Benjamin S. (Ed.). Taxonomy of Educational Objectives: The Classification of Educational Goals. New York: David McKay Company, Inc., 1956.

2. Sanders, Norris M. Classroom Questions: What Kinds? New York: Harper and Row, 1966.

3. Tyler, Ralph W. "Permanence of Learning," Journal of Higher Education, 4:203-204, 1933.